Method of determining the copy number of a nucleotide sequence

The present invention relates to a method of determining the copy number of a nucleotide sequence I in a sample using an amplification technique, said method comprising the steps of

5   1) adding nucleotides, primers, polymerase and any further reagents, if any, required for the amplification technique used to the sample,

2) performing one or more amplification cycles to amplify the nucleotide sequence· I for which the copy number has to be

10   determined;

where the sample contains a chromosomal second nucleotide sequence II, and

a) the first nucleotide sequence I is amplified,

b) the second nucleotide sequence II is amplified,

15   c) a third nucleotide sequence I' corresponding to the first nucleotide sequence I and present in a control sample is amplified at various dilutions, and

d) a fourth nucleotide sequence II' corresponding to·the second nucleotide sequence II and present in a control sample

20   is amplified at various dilutions,

where the ratio of the concentrations of nucleotide sequence I' and II' is known

where the amplifications of the third and fourth nucleotide sequences I' and II' at various dilutions allows standard

25   curves SC_i with i being I or II to be made, the concentrations of I and II are determined by using the respective standard curve SC_i , and the relative concentrations allows the relative copy number CN of sequence I (versus nucleotide sequence II) to be determined using the formula

30

$$CN = \frac{[I]_{SC_{I'}}}{[II]_{SC_{II'}}}$$

35   where

CN is the relative copy number of I over II in the sample;
$[I]_{SC_{I'}}$ is the concentration of I determined using standard
curve $SC_{I'}$; and

1) $[II]_{SC_{II'}}$ is the concentration of II determined using stan-
5   dard curve $SC_{II'}$.

Most eukaryotic diploid cells contain two copies of
a single gene; one on each chromosome of a pair of chromo-
somes. The chromosomes of a pair of chromosomes being derived
from each parent, the genes may be different and, for exam-
10   ple, one of them may result in a abnormal protein. Thus, the
number of functional genes is not necessarily 2 in an eukary-
ote, and can be 1 or even 0. While often genes are present in
one copy per chromosome of a particular pair of chromosomes,
some genes are present in multiple copies, for example in
15   tandem repeat sequences. Another exception to the general
rule of 2 copies per cel is mitochondrial DNA. A cell con-
tains many mitochondria, the number being dependant on the
type of cell. But even for a particular cell type, the number
of mitochondria may vary. Typical numbers are between 100 and
20   1000 mitochondria per cell, and each mitochondrion contains
several copies of mitochondrial DNA. In addition, the typical
copy number is not necessarily equal to larger than 2 per
cell. Some nucleotide sequences are very rare among cells
(despite being of one and the same subject, such as a human
25   being). This is, for example, after gene rearrangement. This
is, for example, the case with antibody producing cells (B-
lymphocytes) or receptor-carrying T-lymphocytes. Of a large
number of lymphocytes, only a few will contain a particular
nucleotide sequence defining the variable region of a par-
30   ticular antibody (or of the T-cell receptor), capable of rec-
ognizing a particular antigen. In the art, a need exists to
reliably determine the copy number of a nucleotide sequence,
which may comprise the nucleotide sequence of a gene or part
thereof. A method according to the preamble is known in the
35   art.

A method according to the preambule is known dis-
closed by Kwok et al in US 5,389,512.

The object of the present invention is to improve

this method for reliably determining the copy number of a nucleotide sequence even if it is present in extreme amounts, such as lots of copies per cell or only few copies per many cells. In addition, an object of the present invention is to
5 provide a method which has reduced sensitivity to the efficiency with which DNA was extracted from the cells containing a nucleotide sequence I for which the copy number has to be determined.

To this end, the method according to the present in-
10 vention is characterized in that at least one pair of amplification reactions chosen from i) a) and b), and ii) c) and d) is performed in a single container and monitored spectrophotometrically during amplification.

This allows for a more accurate measurement of rela-
15 tive or absolute copy numbers of nucleotide sequence I. Suitable spectrophotometrical methods are known in the art. More specifically, such methods rely on internal probes for real time measurements, for example real time PCR. Internal probes are known in the art, and are disclosed by, for example, Wi-
20 ner et al (Anal. Biochem 270, pp. 41-49 (1999)). Measurements can be done either continuously, or after finishing an amplification cycle. While specific reference is made to standard curves, it goes without saying that this can be done using computational methods without an actual graph being made.
25 Hence, in the present application the phrase "making a standard curve" involves any method using at least two reference points to determine a (relative) concentration. Generally, all amplifications will be performed substantially at the same time. By performing multiple amplifications in one con-
30 tainer, the room for error is reduced. The method according to the invention is not only highly accurate, but it is also very efficient if performed for multiple samples. That is, for each nucleotide sequence I for which it is desired to determine the copy number, only a single standard curve $SC_{II'}$
35 has to be made. With respect to the term "corresponding" as used in the present invention in conjunction with nucleotide sequences, this is intended to mean that the nucleotide sequences I and I' (and II and II'), or more specifically the

nucleotide sequence of one and the complementary sequence of the other, are capable of hybridizing under stringent conditions. If the sequences I and I' (and II and II') do not have the same length, the shortest of the two is preferably at
5 most 50% shorter, more preferably at most 30% shorter.

The number of amplification cycles are not necessarily the same for I and II, but they are the same for a) I and I' (both n times); and for b) II and II' (both m times), where n and m are the respective number of amplifications,
10 and hence integers, and n and m may or may not be equal.

Douek et al (Nature 396, pp. 690-695 (1998)) describe a method for detecting the products of the rearrangements of T-cell receptors (TREC) using a semi-quantitative assay. For determining the amount of TREC in a given sample,
15 a known amount of a DNA competitor are prepared. Then, an amount of sample DNA containing the nucleotide sequence to be determined are added to the tube. A PCR amplification reaction is carried out in the presence of radiolabeled deoxynucleotide. Subsequently, the resulting amplification products
20 are run on a gel to separate the sample DNA PCR product from the competitor DNA product. After autoradiography, the amount of nucleotide sequence to be determined is calculated using densitometric analysis from the ratio between a band of competitor DNA and a band of the sample DNA. The result is ex-
25 pressed as the number of copies of TREC per microgram total DNA. To achieve an acceptable accuracy, 4 tubes containing scalar amounts of competitor DNA are used, to which fixed amounts of sample DNA are added. The disadvantage of this method is that when DNA is extracted from cells, it must be
30 assumed that this is all the DNA present in the cells. That is, it is assumed that no cell escaped lysis and all DNA present in the cells was extracted and isolated. This is not necessarily the case. Another disadvantage of this method is that it is sensitive to differences in amplification effi-
35 ciency.

According to a preferred embodiment the absolute copy number is determined by multiplying the copy number CN by the absolute copy number of sequence II per cell.

For several nucleotide sequences II the number of copies of per cell is known. An example is, for example, the gene coding for heat shock protein 70, or Fas Ligand (CD178), which are known to be present with two copies per cell (i.e.

5 the absolute copy number of hsp 70 = 2). Many nucleotide sequences of genes are very suitable because they generally are present in a known number of copies in every cell of the species from which the DNA is derived. The efficiency with which DNA material is extracted from the cells is not important

10 (although, in case nucleotide sequence I is on a different molecule as nucleotide sequence II, it is important that they are extracted with the same efficiency). Hence, this embodiment allows determination of the absolute copynumber of the nucleotide sequence I per cell.

15 According to a preferred embodiment, at least one of the third nucleotide sequence I' and fourth nucleotide sequence II' resides on a vector.

In the present application, a vector is understood to be any nucleotide sequence consisting of or containing the

20 nucleotide sequence(s) to be amplified. When present on a vector capable of being replicated in vitro or in vivo, it is easy to obtain that particular nucleotide sequence in desired quantities. It is also very easy to determine the DNA concentration and hence the copy number of the nucleotide sequence

25 per volume. A vector capable of replication or being replicated may be any such vector known in the art, such as a plasmid, a cosmid, a virus etc. If, according to a favourable embodiment, the third nucleotide sequence I' resides on first vector and the fourth nucleotide sequence II' resides on a

30 second vector, the vectors can be used (or mixed) at any desired ratio to accommodate expected differences in copy number in the sample.

It is highly prefered that the third nucleotide sequence I' and fourth nucleotide sequence II' reside on the

35 same vector.

Thus the ratio is constant and exactly known (for example 1:1). This allows for the most accurate measurements possible.

It is possible to subject the vector containing both nucleo-
tide sequence I' and II' to a digestion using one or more re-
striction enzymes, optionally followed by purification, to
yield a linear molecule containing both both nucleotide se-
5   quence I' and II', and using this molecule for the amplifica-
tions required for the standard curves.

According to a preferred embodiment, at least two
different third nucleotide sequences I' for measuring a cor-
responding number of different first nucleotide sequences I
10 reside on a single vector.

In other words, a single vector, requiring its con-
centration to be determined only once, can carry multiple
third nucleotide sequences I', which allows, for example, the
copy numbers of many different genes to be determined.

15   Preferrably, the sequence of the first nucleotide
sequence I is the same as the third nucleotide sequence I'.

This strongly reduces errors due to differences in
amplification efficiencies between I and I'. Nevertheless,
small differences in nucleotide sequence are generally al-
20 lowed, although changes at locations where the probe used for
detecting the concentration of the nucleotide sequence are
best avoided. In other word, it is highly preferred if the
probe is a perfect match for the sequence where it is in-
tended to bind.

25   Similarly, it is preferred that the sequence of the
second nucleotide sequence II is the same as the fourth nu-
cleotide sequence II'.

While the present invention is described with refer-
ence to DNA, the present invention also applies to the deter-
30 mination of the number of RNA sequences present in a cell.
Use can be made of methods known in the art to multiply RNA,
for example by preparing cDNA. This application does not at-
tempt to teach an interested layman how to become a person
skilled in the art, for which reason the layman is referred
35 to general text books and in particular to a proper univer-
sity to learn the required techniques that a person skilled
in the art knows how to apply these techniques to work the
present invention.

The present invention will now be illustrated with reference to the drawings where

Fig. 1 represents a standard curve for an mtDNA sequence I' (circles) plus data for nucleotide sequence I (squares);

5 Fig. 2 represents a standard curve for a nuclear DNA sequence II' (circles) plus data for nucleotide sequence II (squares);

Fig. 3 represents a standard curve for a nuclear DNA sequence I' (circles) plus data for nucleotide sequence I
10 (squares);

Fig. 4 represents a standard curve for a nuclear DNA sequence II' (FasL) (circles) plus data for nucleotide sequence II (squares); and

Fig. 5 shows the effect of age on the numbers of copies
15 of TREC in peripheral lymphocytes (percentage of lymphocyte cells expressing TREC).

The method according to the invention will be illustrated using two Examples. The first relates to the quantitive analysis of mitochondrial DNA (mtDNA) and demonstrates
20 the technique for determining multiple copies per cell. The second Example demonstrates the quantitative determination of a fractional copy number of a particular nucleotide sequence per cell.

EXAMPLE 1
25 MATERIALS AND METHODS

Primers

The nucleotide sequence I (mtDNA) was a stretch having a length of 102 nucleotides, and corresponds to part of the enzyme NADH dehydrogenase as coded for by mtDNA. Am-
30 plification of nucleotide sequence I was performed using a set of primers, each having a length of 21 nucleotides and synthesized using standard procedures. The sequences of both primers were checked to be unique for human mtDNA using Blast software, through the NCBI site at NIH
35 (http://www.ncbi.nlm.nih.gov/blast/).

The nucleotide sequence II (nuclear DNA) serving as a reference, was a stretch having a length of 104 nucleotides and part of the FasL gene, which comes with two copies per

human cell. Amplification of nucleotide sequence II was per-
formed using a set of primers, each having a length of 21 and
24 nucleotides respectively.

Probes

5   To monitor the progress of amplification, a probe was
used for nucleotide sequence I, the probe having a length of
23 nucleotides, having a FAM (carboxy fluorescein) fluores-
cent probe at the 5' end and a BlackHole Quencher1 ™ group at
the 3' end. This probe, and all others in this application,
10 was ordered commercially with MWG, Ebersberg, Germany. The
sequence of the probe was checked to be unique for human
mtDNA using Blast software, through the NCBI site as men-
tioned above.

The probe used for nucleotide II had a length of 22 nu-
15 cleotides and contained TexasRed as the fluorescent label and
and a BlackHole Quencher2 ™ group at the 3' end (MWG).

DNA isolation

DNA was isolated from HL60, a promyelocytic leukaemia
cell line, using a DNA isolation kit from Qiagen, Hilden,
20 Germany according to the instructions of the manufacturer.

Control

A vector was constructed, using pGEM-11Z (Promega) con-
taining the sequences I' and II' head to tail, using standard
genetic engineering techniques, as all too familiar from Sam-
25 brook et al. (Molecular cloning. A lab manual. (1989)) in E.
coli. The nucleotide sequences I' and II' were identical to
their respective I and II counterparts, and present on the
vector in a highly defined 1:1 ratio.

The absolute concentration of the controls was done us-
30 ing limiting dilution assays (Sambrook).

Amplification

Amplification was performed using an iCycler Thermal cy-
cler (BioRad, Hercules, CA, USA) using standard procedures.
The amplification is performed in plates having 96 wells.
35 This instrument allows monitoring of fluoresence in up to 4
different channels. In short, one cycle of denaturation (95
°C for 6 min) was performed, followed by 45 cycles of ampli-
fication (94 °C for 30 s, 60 °C for 60 s). The amplification

was performed in a mix that consisted of: Promega PCR buffer
1X (Promega, Madison, WI, USA), 3.0 mM MgCl2, 400 pmol of
primers for mtDNA, 0.2 mM dNTP and 2 U of *Taq* polymerase
(Promega). In accordance with the invention, the amplifica-
5 tion for both nucleotide sequences I and II were performed in
a single well, and the same is true for nucleotide sequences
I' and II' (for determining the standard curves). Data were
analysed using the software of the iCycler.

The standard curves were made by introducing a known
10 number of copies of vector per well.

Amplification experiments were performed in triplicate.
RESULTS

Fig. 1 shows the standard curve for nucleotide sequence
I' and Fig. 2 shows the standard curve for the nucleotide se-
15 quence II' based on FasL. Note the excellent correlation co-
efficients of 0.995 and 0.996 respectively, indicating the
excellent accuracy of the method according to the invention.
Using these curves, the concentration of nucleotide sequences
I and II (shown as squares in Figs. 1 and 2) were determined.
20 As it is known that the nucleotide sequence for FasL (and
more specifically for the probe for nucleotide II/II') is
present with two copies per cell, the number of copies of nu-
cleotide sequence I per cell is twice as high, i.e. 160.

25   EXAMPLE 2
Basically, the same method was used as described in Example
1, except that the nucleotide sequence I corresponded to part
of the sequence of the delta locus of the T-cell receptor.
The method was used to determine the number of copies of TREC
30 per cell, in particular peripheral lymphocytes in blood, in
three age groups (healthy humans of 20, 60 or 100 years. The
number of people were respectively 16 (10), 17 (10), and 21
(17), with the number of women between parentheses)

The standard curves for nucleotide sequence I' and II'
35 are shown in fig. 3 and 4 respectively. The following corre-
lation coefficients obtained were: 0.999 and 0.998.

Fig. 5 shows that the number of copies of TREC decreases
with age (averages per age group shown as a horizontal line)

from about 3.2 to 0.1 per 100 cells.

    While particularly beneficial for the method according to the present invention in view of the fact that spectrophotometrical methods allow simulaneous detection of
5 multiple labels, it is possible to perform an amplification reaction using any known amplification technique, where the third nucleotide sequence I' and fourth nucleotide sequence II' resides on a single vector and the amplifications of each of I' and II' are performed in separate containers, such as
10 separate wells. The application covers this possibility as well. Such amplification techniques comprise, apart from the ones mentioned above, CP (Cycling Probe Reaction), bDNA (Branched DNA amplification), SSR (Self-Sustained Sequence Replication), SOA (Strand Displacement Amplification), QBR
15 (Q-Beta Replicase), Re-AMP (Formerly RAMP), NASBA (Nucleic Acid Sequence Based Amplification), RCR (Repair Chain Reaction), LCR (Ligase Chain Reaction), TAS (Transorption Based Amplification System), and HCS (amplifies ribosomal RNA).

CLAIMS

1. Method of determining the copy number of a nu-
cleotide sequence I in a sample using an amplification tech-
nique, said method comprising the steps of

3) adding nucleotides, primers, polymerase and any further
5   reagents, if any, required for the amplification technique
      used to the sample,

4) performing one or more amplification cycles to amplify the
   nucleotide sequence I for which the copy number has to be
   determined;

10 where the sample contains a chromosomal second nucleotide se-
quence II, and

e) the first nucleotide sequence I is amplified,

f) the second nucleotide sequence II is amplified,

g) a third nucleotide sequence I' corresponding to the first
15   nucleotide sequence I and present in a control sample is
      amplified at various dilutions, and

h) a fourth nucleotide sequence II' corresponding to the sec-
   ond nucleotide sequence II and present in a control sample
   is amplified at various dilutions,

20 where the ratio of the concentrations of nucleotide sequence
I' and II' is known
where the amplifications of the third and fourth nucleotide
sequences I' and II' at various dilutions allows standard
curves SC_i with i being I or II to be made, the concentrations
25 of I and II are determined by using the respective standard
curve SC_i , and the relative concentrations allows the rela-
tive copy number CN of sequence I (versus nucleotide sequence
II) to be determined using the formula

30        $$CN = \frac{[I]_{SC_{I'}}}{[II]_{SC_{II'}}}$$

where

35 CN is the relative copy number of I over II in the sample;

$[I]_{SC_{I'}}$ is the concentration of I determined using standard curve $SC_{I'}$; and

$[II]_{SC_{II'}}$ is the concentration of II determined using standard curve $SC_{II'}$

5   **characterized** in that at least one pair of amplification reactions chosen from i) a) and b), and ii) c) and d) is performed in a single container and monitored spectrophotometrically during amplification.

      2. Method according to claim 1, **characterized** in
10   that the absolute copy number is determined by multiplying the copy number CN by the absolute copy number of sequence II per cell.

      3. Method according to claim 1 or 2, **characterized** in that the third nucleotide sequence I' and fourth nucleo-
15   tide sequence II' resides on a single vector.

      4. Method according to claim 3, **characterized** in that at least two different third nucleotide sequences I' for measuring a corresponding number of different first nucleotide sequences I reside on a single vector.

20   5. Method according to any of the preceding claims, **characterized** in that the sequence of the first nucleotide sequence I is the same as the third nucleotide sequence I'.

      6. Method according to any of the preceding claims, **characterized** in that the sequence of the second nucleotide
25   sequence II is the same as the fourth nucleotide sequence II'.

FIG.1

Correlation Coefficient: 0.995 Slope: -3.329 Intercept: 35.314 Y=-3.329 X+35.314
PCR Efficiency: 99.7%

□ Unknowns
◇ Standards

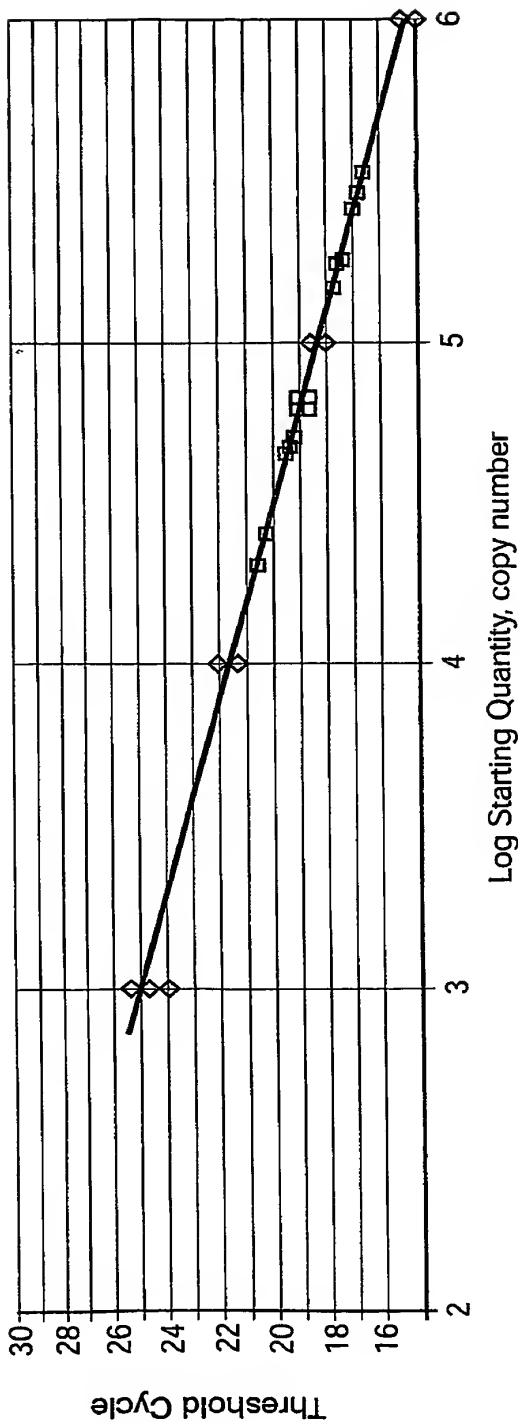Log Starting Quantity, copy number

Threshold Cycle

FIG.2

FIG.3

Correlation Coefficient: 0.998 Slope: -3.611 Intercept: 37.621 Y=-3.611 X+37.621

FIG.4

Fig. 5